Optical detection and monitoring of pigmented skin lesions

J. J. STAMNES,[1,2,*] **G. RYZHIKOV,**[1] **M. BIRYULINA,**[1] **B. HAMRE,** [1,2] **L. ZHAO,**[1] **AND K. STAMNES,**[1,3]

[1]*Balter Medical AS, Thormølensgate 55, 5008 Bergen, Norway*
[2]*Department of Physics and Technology, University of Bergen, P.O. Box 7803, 5020 Bergen, Norway*
[3]*Department of Physics and Engineering Physics, Stevens Institute of Technology, Hoboken, NJ, USA*
[*]*jakob.stamnes@uib.no*

Abstract: A method is presented for discriminating between malignant and benign pigmented skin lesions based on multispectral and multi-angle images. It is discussed how to retrieve maps of physiology properties and morphometric parameters from recorded images using a bio-optical model, radiative transfer calculations, and nonlinear inversion, and how to employ automated zooming to extract lesion and surrounding masks. Training and validation of a classification scheme for separation between benign and malignant tissue yielded sensitivity/specificity ranging from 97%/97% for application to a small dataset comprised of lesions not used for training and validation to 99%/93% for application to a larger dataset.

References and links

1. V. Foerster, "Optical scanners for melanoma detection" [Issues in emerging health technologies, Issue 123]. Ottawa: Canadian Agency for Drugs and Technologies in Health (2014).
2. E. A. Kupetsky and L. K. Ferris, "The diagnostic evaluation of MelaFind multi-spectral objective computer vision system," Expert Opin. Med. Diagn. **7**, 405–411 (2013).
3. S. J. Divito and L. K. Ferris, "Advances and short comings in the early diagnosis of melanoma," Melanoma Res. **20**, 450–458 (2010).
4. C. Herman, "Emerging technologies for the detection of melanoma: achieving better outcomes," Clin. Cosmet. Investig. Dermatol. **5**, 195–212 (2012).
5. H. Lui, J. Zhao, D. McLean, and H. Zeng, "Real-time Raman spectroscopy for in vivo skin cancer diagnosis," Cancer Res. **72**, 2491–2500 (2012).
6. Verisante AuraTM [Internet], "Vancouver: Verisante Technology," 2014 [cited 2014 Feb 21]. Available from: http://www.verisante.com/products/aura/
7. P. Fayerman, "Vancouver skin cancer detection device ready for market: Nobel prize-winning research behind it" [Internet]. Vancouver: Vancouver Sun; 2013 Dec 14. [cited 2014 Feb 20]. Available from: http://blogs.vancouversun.com/2013/12/14/vancou ver-skin-cancer-detection-device-ready-for- market-nobel-prize-winning-research-behind-it/
8. G. Monheit, A. B. Cognetta, L. Ferris, H. Rabinovitz, K. Gross, M. Martini, J. M. Grichnik, M. Mihm, V. G. Prieto, P. Googe, R. King, A. Toledano, N. Kabelev, M. Wojton, and D. Gutkowicz-Krusin, "The performance of MelaFind: a prospective multicenter study," Arch. Dermatol. **147**, 188–194 (2011).
9. MelaFind [Internet]. Irvington (NY): MELA Sciences; 2014. What is MelaFind; 2014 [cited 2014 Feb 20]. Available from: http://www.melafind.com/patients/about-melafind
10. SIMSYSTM MoleMateTM [Internet]. Mississauga (ON): MedX Health Corporation. 2014 [cited 2014 Feb 20]. Available from: http://simsys- molemate.com/
11. F. M. Walter, H. C. Morris, E. Humphrys, P. N. Hall, A. T. Prevost, N. Burrows, L. Bradshaw, E. C. F. Wilson, P. Norris, J. Walls, M. Johnson, A. L. Kinmouth, and J. D. Emery, "Effect of adding a diagnostic aid to best practice to manage suspicious pigmented lesions in primary care: randomised controlled trial," BMJ, 345:e4110 (2012).
12. G. Pellacani, A. M. Cesinari, and S. Seidenai, "Reflectance mode confocal microscopy of pigmented skin lesions – improvement ooin melanoma diagnostic specificity," J. Am. Acad. Dermatol. **53**, 979–985 (2005).
13. P. Guitera, G. Pellacani, C. Longo, S. Seidenari, M. Avramidis, and S. W. Menzies, "In vivo reflectance confocal microscopy enhances secondary evaluation of melanocytic lesions," J. Invest. Dermatol. **129**, 131–138 (2009).
14. D. Wielowieyska-Szybinska, K. Biatek-Gals, K. Podolec, and A. Wojas-Pelc, "The use of reflectance confocal microscopy for examination of benign and malignant skin tumors," Postep. Derm. Alergol. **6**, 380-387 (2014).

Research Article   Vol. 8, No. 6 | 1 Jun 2017 | BIOMEDICAL OPTICS EXPRESS 2947

Biomedical Optics EXPRESS

15. D. L. Swanson, S. Laman, M. S. Biryulina, K. P. Nielsen, G. A. Ryzhikov, J. J. Stamnes, B. Hamre, L. Zhao, F. S. Castellana, and K. Stamnes, "Optical transfer diagnosis of pigmented lesions: a pilot study," Skin Res. Tech. **15**, 330–337 (2009).
16. D. L. Swanson, S. D. Laman, M. S. Biryulina, K. P. Nielsen, G. A. Ryzhikov, J. J. Stamnes, B. Hamre, L. Zhao, E. R. Sommersten. F. S. Castellana, and K. Stamnes, "Optical Transfer Diagnosis of Pigmented Lesions," Dermatol. Surg. **36**, 1979–1986 (2010).
17. K. P. Nielsen, L. Zhao, G. A. Ryzhikov, M. S. Biryulina, E. R. Sommersten, J. J. Stamnes, K. Stamnes, and J. Moan, "Retrieval of the physiological state of human skin from UV-Vis reflectance spectra: A feasibility study," Photochem. Photobiol. B **93**, 23–31 (2008).
18. K. P. Nielsen, L. Zhao, P. Juzenas, J. J. Stamnes, K. Stamnes, and J. Moan, "Reflectance spectra of pigmented and non-pigmented skin in the UV spectral region," Photochem. Photobiol. **80**, 450–455 (2004).
19. D. L. Swanson, R. V. Venneugues, S. Q. Vicensio, J. Garioch, M. S. Biryulina, G. A. Ryzhikov, B. Hamre, L. Zhao, F. S. Castellana, K. Stamnes, and J. J. Stamnes, "Optical transfer diagnosis differentiating benign and malignant pigmented lesions in simulated primary care practice," to be submitted, 2017.
20. J. Malvehy, A. Hauschild, C. Curiel-Lewandrowski, P. Mohr, R. Hofmann-Wellenhof, R. Motley, C. Berking, D. Grossmann, J. Paoli, C. Loquai, J. Olah, U. Reinhold, H. Wenger, T. Dirschka, S. Davis, C. Henderson, H. Rabinovitz, J. Welzel, D. Schadendorf, and U. Birgersson, "Clinical performance of the Nevisense system in cutaneous melanoma detection: an international, multicentre, prospective and blinded clinical trial on efficacy and safety," Br. J. Dermatology **171**, 1099–1107 (2014).

1. Introduction

In recent years, considerable effort has been expended on devoloping optical methods for characterizing tissue and monitoring changes. In 2014, the Canadian Agency for Drugs and Technologies in Health published a report [1] on optical scanners for melanoma detection, discussing three different devices (Aura, MelaFind, and SIMSYS-MoleMate) approved for marketing in Canada and/or USA. As the report confirms, the 5-year survival rate is 93 to 97% for melanoma detected at an early stage, but drops to between 10 and 20% for advanced stage detection, implying that there is a need for accurate diagnostic devices to enable early detection while avoiding unnecessary biopsies [2–5].

Aura (Verisante Technology, Inc., Vancouver, British Columbia, Canada) utilizes near-infrared laser light and Raman spectroscopy to distinguish malignant from benign skin lesions [6, 7]. A clinical study [5], involving 453 patients with 518 suspicious pigmented lesions, of the diagnostic performance for discrimination between melanoma and benign pigmented lesions showed sensitivities ranging from 90 to 99% for specificities ranging from 68 to 15%.

MelaFind (MELA Sciences, Inc., Irvington, New York, USA) illuminates the skin at 10 wavelengths, measures light scattered back from the skin, and uses image analysis algorithms combined with a skin disorder database to provide treatment suggestion [8, 9]. For discrimination between melanoma and benign pigmented lesions in a population of suspicious lesions a clinical study involving 1,383 patients with 1,831 pigmented skin lesions showed 98% sensitivity and 9% specificity [8].

SIMSYS-MoleMate Skin Imaging System (MedX Health, Inc., Hamilton, Ontario, Canada) is based on using a handheld, multispectral scanner and computer software to provide dermoscopic images, dermal and epidermal pathological characteristics, and the ability to catalogue, monitor, and compare lesions over time [10]. In a randomized controlled trial involving 1,297 patients with 1,580 suspicious pigmented lesions it was found that adding MoleMate to best practice resulted in lower agreement with expert assessment that the lesion was benign and led to a higher proportion of referrals [11].

An algorithm based on reflectance confocal microscopy (RCM) has been proposed [12] for diagnosis of melanoma, and its sensitivity and specificity has been compared to that of dermoscopy [13], showing RCM to have specificity of 68% compared to 32% for dermoscopy, whereas the sensitivity was 91% for RCM and 88% for dermoscopy [14].

In two previous studies [15, 16], the application of optical transfer diagnosis (OTD) for non-invasive detection of skin cancer was discussed, the objective being to evaluate the potential

of this technology for differentiation of benign from malignant pigmented melanocytic lesions. After scanning, lesions were biopsied for histopathologic examination, each by two separate dermatopathologists. The second study [16] included 94 patients with a total of 118 pigmented lesions suggestive of melanoma, out of which 11 were identified as melanoma or atypical melanocytic hyperplasia consistent with melanoma. For identification of melanomas, OTD had a sensitivity of 100% and a specificity of 90%. The results in [15, 16] were based on using a linear binary classifier to differentiate between malignant and benign lesions in a population of suspicious lesions. As the set of measurements on suspicious lesions was enlarged with similar types of lesions beyond the 118 lesions studied in [16], a linear binary classifier was found not to perform satisfactorily, probably due to overtraining on the limited initial data set, providing overoptimistic results that could not be sustained as the data set was enlarged. The objective of the new classification scheme discussed in this paper, which involves clustering and construction of nonlinear classifiers, is to discriminate between two classes of tissue, e.g. malignant vs. benign. For application to pigmented unselected ("all-comer") skin lesions, we demonstrate that such discrimination can be performed with high sensitivity and specificity.

In addition to clustering, the new classification scheme includes the following novel features: (i) estimation of noise using multiple measurements (typically three) of each lesion to improve data quality and increase robustness (see Appendix A); (ii) use of several new physiology-based diagnostic parameters to improve separation of lesion types; (iii) use of generalized diagnostic parameters to reduce their number; and (iv) a new cost function to increase robustness.

The objective of this paper is to describe new features of the OTD classification scheme compared to those presented in [15, 16], and present results that demonstrate its potential.

2.   Optical transfer diagnosis

An optical transfer diagnosis (OTD) device (Balter Medical AS, Bergen, Norway) is a spectral radiometer that records a set of 30 images, constituting a lesion measurement, in less than 10 seconds. Images are recorded at 10 different wavelengths (365–1,000 nm) from multiple angles of illumination and detection.

The OTD system consists of a handheld unit, a docking station, and a laptop PC, as shown in Fig. 1. The only control item on the handheld unit is a release button to initiate image recording, all other controls are performed via a user interface on the laptop PC. When not in use, the handheld unit is placed in the docking station, where the observation window is placed against a calibration target. Figure 2 shows the inner parts of the handheld OTD unit. The sensor head
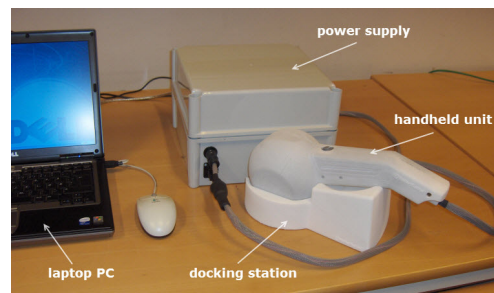

Fig. 1. Illustration of the OTD system with its handheld unit placed in a docking station, and with cables connecting it to a power supply and a laptop PC.

contains an illuminating system and an imaging system, as shown in the three-dimensional sketch in Fig. 3. The illuminating system consists of 12 fixed light-emitting diode (LED) lamps. Each LED is placed at a different angle relative to the skin to enhance the ability to retrieve depth information. The polar angles of the LEDs vary between 30 and 45°, and the relative
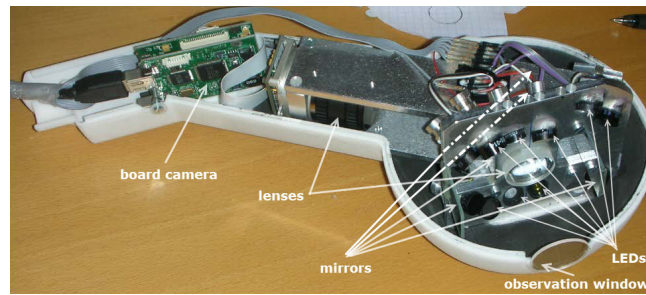
Fig. 2. Illustration of the inner parts of the handheld OTD unit.

azimuth angles between 34 and 145°. The polar angles for the detectors vary between 0 and 45°, and the relative azimuth angles between 0 and 180°.

The imaging system consists of one correcting lens placed inside the handle plus another correcting lens and five mirrors placed inside the sensor head (Figs. 2, 3, and 4) and a sapphire glass observation window that contacts the area of the skin lesion. The camera chip consists of three IEEE (Institute of Electrical and Electronics Engineers) 1394 FireWire cameras.
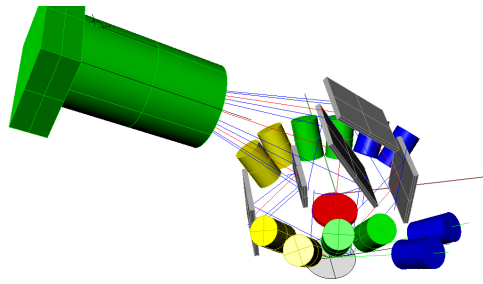

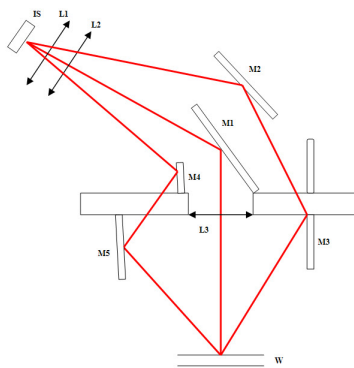Fig. 3. 3D sketch of the inner parts of the handheld OTD unit.

Fig. 4. Ray trace of the OTD imaging system. IS: Image sensor: L1: Camera lens; L2: Correcting lens 1; L3: Correcting lens 2; M1: Plane mirror for central view image; M2, M3: Plane mirrors for upper 30° oblique view image; M4, M5: Plane mirrors for lower 45° oblique view image; W: Sapphire glass window.

As indicated in Fig. 4, the five mirrors are used to image the same area of the skin viewed

from three different angles on three different sections of the camera chip. Figure 4 shows the central rays for images from three different view angles. The plane mirror M1 provides a nadir (vertically downward) view image, the plane mirrors M2 and M3 give an upper 30° oblique view image, and the plane mirrors M4 and M5 provide a lower 45° oblique view image. To compensate for different object distances for the three angular views, the camera chip is slightly tilted relative to the optical axis. Also, as explained in section 3, image compression is employed, which reduces the resolution requirement significantly.

An alcohol-based gel interface is used where the sapphire observation window contacts the skin to provide refractive-index matching and avoid rough-surface scattering, and to obtain illumination and imaging of a selected area of the skin through the 2.2-cm-diameter circular sapphire observation window.

On the basis of established absorption and transmission spectra for known skin chromophores and mathematical modeling of skin reflectance [17, 18], the set of recorded images is used to retrieve maps of physiology properties and morphometric parameters of the lesion, which are assumed to be different for benign and malignant tissue. The retrieval is based on (i) a bio-optical model that relates physiological properties of skin tissue to its inherent optical properties [18], (ii) a radiative transfer model that for a given set of inherent optical properties computes the light backscattered from the skin for a given wavelength and direction of illumination, and (iii) a nonlinear inversion algorithm that compares the computed backscattered light at various wavelengths and directions with that of the recorded image set [17].

The data acquisition geometry is designed in such a way that for each combination of illumination and detection directions the same area of the skin is interrogated, allowing a one-dimensional treatment when the independent-column approximation is invoked and the skin tissue is assumed to have a layered structure: an uppermost layer, the *epidermis*, consisting of an upper part and a lower part; the *dermis*, containing the blood circulation; and the *subcutis*, a strongly scattering fat-containing layer. The inherent optical properties of each layer are the absorption and scattering coefficients as well as the scattering phase function (describing the angular variation of the scattered light), each varying with wavelength. The retrieved physiology properties and morphometric parameters are (1) percentage of hemoglobin concentration, (2) percentage of hemoglobin oxygenation, (3) upper epidermal thickness, (4) lower epidermal thickness, (5) percentage of melanosome concentration in upper epidermis, (6) percentage of melanosome concentration in lower epidermis, and (7) percentage of keratin concentration. Each of these seven physiology properties or morphometric parameters is retrieved pixel by pixel in the compressed image to provide a map covering the zoomed lesion area. Note that there is no verification of any of these retrieved values by application of different methods for measuring them, but they have been found useful in an intermediate step leading to a final diagnostic indication, as explained below.

From each map, an entropy value is calculated and cross entropy values are calculated for different pairs of maps. The entropy concept used here is similar to that used in statistical physics and information theory. Thus, we define the entropy $E(p_i)$ associated with map number i ($i = 1, 2, 3, \ldots, 7$) in terms of the spatial distribution $p_i(\mathbf{r})$ of the relevant physiological property or morphometric parameter, where $\mathbf{r}$ is the position inside the area Ω of the map, as follows

$$E(p_i) = - \int_\Omega \lambda_i p_i(\mathbf{r}) \ln[\lambda_i p_i(\mathbf{r})] d\mathbf{r} \; ; \; \lambda_i = \left[\int_\Omega p_i(\mathbf{r}) d\mathbf{r} \right]^{-1} \tag{1}$$

and we define the cross entropy for all pairs of maps with spatial distributions $p_i(\mathbf{r})$ ($i = 1, 2, 3, \ldots, 7$) and $p_j(\mathbf{r})$ ($j = 1, 2, 3, \ldots, 7$) as the symmetic matrix $\mathbf{E}$ with elements E_{ij}, given by

$$E_{ij} = \frac{1}{2} \left[-E(p_i) - E(p_j) + \int_\Omega \lambda_i p_i(\mathbf{r}) \ln[\lambda_j p_j(\mathbf{r})] d\mathbf{r} + \int_\Omega \lambda_j p_j(\mathbf{r}) \ln[\lambda_i p_i(\mathbf{r})] d\mathbf{r} \right]. \tag{2}$$

These entropy and cross entropy values are used to define diagnostic parameters, as discussed in Section 4.

3.   Image recording and pre-processing

A lesion measurement comprises a set of 30 images recorded by an OTD scanner. For a given wavelength and direction of illumination, the OTD scanner records three images simultaneously at different detection angles. This procedure is repeated for 9 other wavelengths in the range from the near ultraviolet to the near infrared at different illumination angles to produce a lesion measurement that comprises a set of 30 images.

For any of the 30 images, each pixel corresponds to (i) a particular distance from one of the 10 LED sources of different intensity, (ii) a particular size of the of skin area for each of the 20 images that are recorded by one of the two side-viewing cameras; (iii) a particular location of the illuminated skin area because of possible movement of the skin with respect to the OTD scanner during the few seconds of sequential illumination by the 10 LEDs.

To address issues (i)-(iii) above, a series of pre-processing steps are performed, including (1) relative calibration such that the intensity of each pixel is measured in units of the intensity due to backscattering from a corresponding pixel having a Lambert surface; (2) geometrical calibration such that an ellipse of illuminated skin area for a side-viewing camera is transformed into a circle; (3) image registration such that each pixel in any of the 30 images corresponds to the same area of the skin; (4) compression such that the raw image having about 7×10^6 pixels with a spatial resolution of about 25 μm is replaced by a smoothed image with a total number of pixels that is 100 times less than that of the raw image. Thus, the compressed image has 10 times less spatial resolution than the raw image, and a by-product of compression is filtering of spatial high-frequency noise in the raw images due to possible spikes, specular points, and hairs.

Automated zooming is employed to provide a lesion mask that circumferences the lesion and is characteristic of its shape, and a surrounding mask that creates an area outside the lesion of suitable size and the same outer shape as that of the lesion mask. The zooming allows one to get rotation, translation, and scaling invariant characteristics of the lesion under investigation, both for calibrated images and maps of physiology properties and morphometric parameters. Also, zooming accelerates the processing since only pixels of the lesion and surrounding masks are considered.

Figure 5 shows a dermoscopic image of a melanoma, while Fig. 6 shows the corresponding RGB image and maps of physiology properties and morphometric parameters obtained from OTD recordings and processing. Figures 7 and 8 show corresponding results for a compound nevus. As noted above, from these maps, entropy and cross entropy values are calculated and used to define diagnostic parameters, as discussed in Section 4.

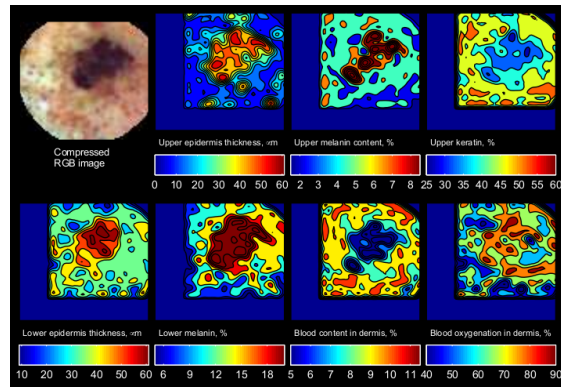Fig. 5. Dermoscopic image of a melanoma.

Fig. 6. Corresponding RGB image and maps of physiology properties and morphometric parameters for the melanoma in Fig. 5 obtained from OTD recordings and processing.
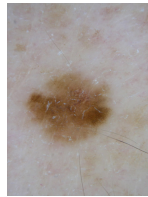

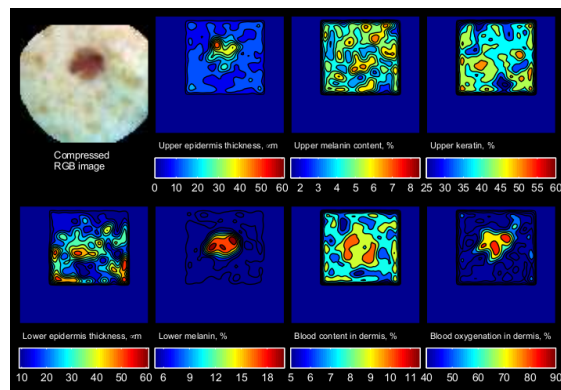Fig. 7. Dermoscopic image of a compound nevus.

Fig. 8. Corresponding RGB image and maps of physiology properties and morphometric parameters for the compound nevus in Fig. 7 obtained from OTD recordings and processing.

4. Diagnostic parameters

From the calibrated, registered, compressed, and zoomed OTD image of a lesion obtained from nadir (vertically downward) illumination by green light (hereafter referred to as the 'nadir green image') the following 10 morphometric parameters are derived: (1) Size; (2) Histogram width (providing a measure of inhomogeneity of the reflected intensity); (3) Fractal dimension; (4) Moment of inertia; (5) Asphericity; (6) Center distance (representing the physical distance between the geometrical center of the lesion and the center of mass of absorptance); (7) Border length; (8) Average darkness; (9) Area divided by fractal dimension; and (10) Border length divided by fractal dimension.

From the seven maps obtained from the retrieval discussed in Section 2, seven entropies and 21 cross entropies are derived, providing a total of 28 physiology properties and morphometric parameters. By including also the logarithm of each of the 10 morphometric parameters obtained from the nadir green image and the 28 entropy and cross entropy values derived from the seven maps, one obtains a total of 76 diagnostic parameters.

Another 10 diagnostic parameters are derived from the maps of physiology properties: (1) Maximum value of the melanin optical depth in the lesion area; (2) Architectural disorder: ratio of maximum to minimum value of the melanin optical depth in the lesion area; (3) Blood filling: maximum value of blood content in the surrounding area; (4) Angiogenesis: ratio of the number of blood vessels in a surrounding area close to the lesion border to that in an area farther from the lesion border, given by $c_1 A_1 \ell_2 / c_2 A_2 \ell_1$, where (with $j = 1, 2$) c_j is the blood concentration in area A_j, and ℓ_j is the distance from the center of the lesion to the outer border of area A_j. Here A_1 is a surrounding area close to the lesion border, and A_2 is a surrounding area farther from the lesion border; (5) Ratio of the blood oxygenation in a surrounding area close to the lesion border to that in an area farther from the lesion border, given by $c'_1 A_1 \ell_2 / c'_2 A_2 \ell_1$, where (with $j = 1, 2$) c'_j is the blood oxygenation in area A_j, and where ℓ_j and A_j are the same as in the definition above of Angiogenesis; (6) Melanin contrast: ratio of the total melanin optical depth in the lesion area to that in the surrounding area; (7) Blood contrast: ratio of the blood content in the lesion area to that in the surrounding area; (8) High spatial Fourier-components of the map of total melanin optical depth in the lesion area (natural RoI, standard gridding); (9) Entropy of contrast of the map of total melanin optical depth in the lesion area (natural RoI, standard gridding); (10) The same entropy of contrast as above, but in the "original RoI".

Here "RoI" stands for "Region of Interest", and "natural RoI" stands for a rectangular area that is oriented in accordance with the shape of the lesion, and "standard gridding" means that along the longest side of the natural RoI there are 100 grid points. The "original RoI" is the rectangular zoom area of the compressed digital image.

As discussed above, there are $N = 86$ diagnostic parameters p_j ($j = 1, 2, \ldots, N$): 2×10 morphometric parameters derived from the nadir green image; 2×28 entropies and cross entropies derived from maps of physiology properties and morphometric parameters; and 10 additional physiology parameters derived from maps of physiology properties.

5. Diagnostic index and clustering

5.1. Diagnostic index

For each independent lesion measurement, we define a *diagnostic index D* as a weighted sum of the diagnostic parameters p_j

$$D = \mathbf{w} \cdot \mathbf{p}. \tag{3}$$

Here the *weight vector* $\mathbf{w}$ consists of N weights w_j ($j = 1, 2, \ldots, N$), and $\mathbf{p}$ is a vector of N diagnostic parameters p_j.

5.2. Clustering

Preliminary results based on analyses of independent measurements performed on a limited set of *suspicious* lesions [15, 16] indicated that a *linear binary* classifier would suffice to discriminate between independent measurements on *malignant* (class 1) lesions and *benign* (class 2) lesions. But as the set of *suspicious* lesions was enlarged with the same type of lesions, a linear binary classifier was found *not* to perform satisfactorily, motivating the development and use of *clustering*, which is used to obtain discrimination between class 1 and class 2 lesions through the identification of one set of class 1 clusters and another set of class 2 clusters, each cluster comprising a certain number of independent measurements. Standard methods of clustering were considered but found not to be applicable since it is not known a priori how the diagnostic parameters p_j in Eq. (3) can be used to discriminate between objects from opposite classes, which are organized in accordance with dermoscopy evaluations and biopsy results.

For training of a diagnostic indication algorithm, we consider a set of lesions, some belonging to class 1 and others to class 2, for each of which the diagnosis is known, and let each independent lesion measurement be characterized by N diagnostic parameters p_j ($j = 1, 2, \ldots, N$). The first step of the clustering procedure is to *discretize* the diagnostic parameter p_j as follows:

1. Calculate its mean value μ_j by averaging over the set of independent measurements on class 2 lesions and its standard deviation σ_j by averaging over the set of independent measurements on class 1 lesions, which has higher variability than the set of measurements on class 2 lesions.

2. Discretize p_j by setting it equal to p_j^*, where

$$p_j^* = \begin{cases} -1 & \text{if } p_j < \mu_j - 0.7\sigma_j \\ 0 & \text{if } \mu_j - 0.7\sigma_j < p_j < \mu_j + 0.7\sigma_j \\ +1 & \text{if } p_j > \mu_j + 0.7\sigma_j \end{cases} \qquad (4)$$

where the cutoff value of $0.7\sigma_j$ is chosen in order to ensure a fair representation of the set of measurements on class 2 lesions. Note that the diagnostic parameters for class 2 (benign) lesions have less variation than those for class 1 (malignant) lesions, implying that the trace of the covariance matrix for class 1 lesions is larger than that for class 2 lesions. Therefore, the mean value above is obtained from class 2 lesions, whereas the standard deviation is obtained from class 1 lesions.

The definition of a clustering index for an independent lesion measurement is based on constructing coincidence vectors $\mathbf{C}^+$ and $\mathbf{C}^-$ and probabilistic vectors $\mathbf{t}^+$ and $\mathbf{t}^-$:

$$\mathbf{C}^\pm = \{T_1^\pm, T_2^\pm, \ldots, T_N^\pm\} \; ; \quad \mathbf{t}^\pm = \{t_1^\pm, t_2^\pm, \ldots, t_N^\pm\} \qquad (5)$$

where the components $T_j^\pm$ are coincidence parameters given by $T_j^+ = 1$ if $p_j^* = +1$ and $T_j^+ = 0$ otherwise; $T_j^- = 1$ if $p_j^* = -1$ and $T_j^- = 0$ otherwise; and where $t_j^\pm = (\sum T_j^\pm)^{1/2}$, the sum being over all independent measurements on lesions of the class under consideration. The clustering index C for an independent measurement on either a class 1 or a class 2 lesion is given by:

$$C = \mathbf{t}^+ \cdot \mathbf{C}^+ + \mathbf{t}^- \cdot \mathbf{C}^-. \qquad (6)$$

The independent measurements are ordered in accordance with the value of the clustering index, and independent measurements having values of the clustering index in a specific interval, are taken to belong to the same cluster. For details, see Appendix B.

6. Generalized diagnostic parameters, cost function and optimization

6.1. Generalized diagnostic parameters

We reduce the number of diagnostic parameters by (i) introducing a covariance matrix for independent measurements on lesions of class 1 and class 2, (ii) defining a discriminating operator (see Eq. (25)) in terms of the two covariance matrices, (iii) constructing eigenvectors and eigenvalues on the basis of the discriminating operator and using only those eigenvalues that are larger than a threshold value, chosen so as to ensure that sufficiently large variations of the diagnostic parameters associated with independent measurements on class 1 lesions are accounted for, and (iv) defining for each independent measurement on a lesion of class 1 or class 2 a set of *generalized* diagnostic parameters. As a result, Eq. (1) becomes

$$D = \tilde{\mathbf{w}} \cdot \tilde{\mathbf{p}} \tag{7}$$

where $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{p}}$ are generalized weight and diagnostic parameter vectors, respectively, each having a dimension $\tilde{N}$ that is typically only one third of the number N of original diagnostic parameters. For details, see Appendix C.

6.2. Cost function and optimization

To determine optimal values of the weights in Eq. (7) we define a cost function consisting of a master term and a constraint term, where the latter is used to constrain the length of the weight vector to lie on the surface of a hypersphere in $\tilde{N}$ dimensions of radius equal to 1. For discussion of the master term of the cost function, consider a set of independent lesion measurements that is divided into two subsets of independent measurements, one on lesions belonging to class 1 (malignant) and another on lesions belonging to class 2 (benign).

For each generalized weight vector $\tilde{\mathbf{w}}$, we compute, using Eq. (7), the corresponding generalized diagnostic indices $D_1(\tilde{\mathbf{w}})$ and $D_2(\tilde{\mathbf{w}})$ for each of the class 1 and class 2 independent lesion measurements, respectively. Next, we compute the mean values $\langle D_1 \rangle$ and $\langle D_2 \rangle$ and the corresponding standard deviations σ_1 and σ_2. The master term of the cost function is given in terms of these parameters as

$$J_0(\tilde{\mathbf{w}}) = \frac{1}{\sigma_1} \int_{D^*(\tilde{\mathbf{w}})}^{\infty} \exp\{-\frac{[D - \langle D_1 \rangle]^2}{2\sigma_1^2}\}dD + \frac{1}{\sigma_2} \int_{-\infty}^{D^*(\tilde{\mathbf{w}})} \exp\{-\frac{[D - \langle D_2 \rangle]^2}{2\sigma_2^2}\}dD \tag{8}$$

where $D^*(\tilde{\mathbf{w}})$ is the point of intersection of the two Gaussians in Eq. (8), and the value of $J_0(\tilde{\mathbf{w}})$ is the area of overlap of the two Gaussians. The minimization of $J_0(\tilde{\mathbf{w}})$ will provide the smallest degree of overlap between the two Gaussians, and hence the best separation between independent measurements on class 1 and class 2 lesions. After minimization of the cost function, Eq. (7) becomes

$$D = \mathbf{e} \cdot \tilde{\mathbf{p}} \tag{9}$$

where $\mathbf{e}$ is an optimal generalized weight vector, hereafter referred as an *expert*, regarding the separation between independent measurements belonging to the two classes of lesions.

7. Partial reliabilities for a pair of opposite clusters

For a *pair of opposite clusters*, consisting of cluster #i of class 1 lesions and cluster #j of class 2 lesions, we obtain a *probabilistic* characterization of an expert by (1) computing for each independent lesion measurement included in the pair, the D value, given by $D = \mathbf{e} \cdot \tilde{\mathbf{p}}$ [see Eq. (9)], where $\mathbf{e}$ is the expert; (2) constructing two histograms, one for each cluster in the pair, where each histogram represents the number of independent measurements having D values within different bins in the interval between the minimum D value ($D_{\min}$) and the maximum

D value ($D_{\max}$). Here $D_{\min}$ is the absolute value of the largest negative D value for class 2 lesions (see blue curve in Fig. 9), and $D_{\max}$ is the largest D value for class 1 lesions (see red curve in Fig. 9). As a result, we obtain the two histograms (red for class 1 and blue for class 2) illustrated by the vertical lines in Fig. 9; (3) constructing the corresponding two smooth histograms in Fig. 9 that represent probability density functions (pdfs); and (4) integrating the blue (class 2) pdf in Fig. 9 from $D_{\min}$ to a given D value and the red (class 1) pdf from a given D value to $D_{\max}$ to obtain the corresponding cumulative distributions in Fig. 10.
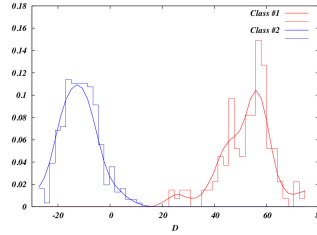

Fig. 9. Two histograms: *class 1* (red) and *class 2* (blue), and corresponding smooth representations.
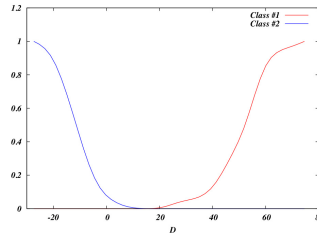

Fig. 10. Cumulative distributions corresponding to the smooth histograms in Figure 1. *Class 1* (red) and *Class 2* (blue).

For an independent lesion measurement having a certain D value, we interpret the corresponding ordinate values r on the two distribution curves in Fig. 10 as *partial reliabilities* of a diagnostic indication. If we let $r_{i,j}^{(1)}(D)$ represent the red (class 1) curve in Fig. 10 and $r_{i,j}^{(2)}(D)$ represent the blue (class 2) curve, then, by definition, $r_{i,j}^{(1)}(D)$ [$r_{i,j}^{(2)}(D)$] represents the partial reliability of a diagnostic indication in favor of the measurement belonging to a class 1 [class 2] lesion.

8. Diagnostic indication by a team of experts

Consider a *randomly drawn* training ensemble comprised of M_1 and M_2 clusters of class 1 and class 2, respectively, and define a *modified* diagnostic index $\tilde{D}$ by

$$\tilde{D} = \text{sign}(\delta_{m_1})|\delta_{m_1}|_{\max}^{1/M} \quad ; \quad \delta_{m_1} = \prod_{m_2=1}^{M_2} (r_{m_1,m_2}^{(1)} + \epsilon) - \prod_{m_2=1}^{M_2} (r_{m_1,m_2}^{(2)} + \epsilon) \qquad (10)$$

where $M = M_1 \times M_2$ and $|\delta_{m_1}|_{\max}$ is the largest of the $|\delta_{m_1}|$ values for $m_1 = 1, 2, \ldots, M_1$. The value $\epsilon = 0.001$ is used to avoid zeros appearing in the products in Eq. (10).

The diagnostic indication by the team of $M = M_1 \times M_2$ experts associated with this randomly drawn training ensemble is that if $\tilde{D}$ given by Eq. (10) is greater than or equal to zero, then the measurement is regarded to represent a class 1 lesion. Typically three measurements are taken of each lesion, and the diagnostic indication for a lesion by the team of experts is that if the mean value of the modified diagnostic indices $\tilde{D}$ given by Eq. (10) for the measurements taken is greater than or equal to zero, the lesion is regarded to be of class 1.

Figure 11 shows a flow chart of the steps followed to obtain a diagnostic indication by a team of experts. (1) Draw at random from a given large set (GLS) of lesions, 77% of the independent measurements (IMs) to obtain a training ensemble (TE); (2) Calculate for each IM in the TE $N = 86$ diagnostic parameters (DPs); (3) Perform clustering to obtain M_1 clusters of class 1 (C1) and M_2 clusters of class 2 (C2); (4) Calculate for each IM in the GLS genneralized diagnostic parameters (GDPs) given by Eq. (7) with an initial generalized weight vector (GWV) of equal components; (5) Minimize a cost function (CF) to obtain $M_1 \times M_2$ optimal GWVs or *experts*, one for each opposite cluster (OC) pair; (6) Calculate partial reliabilities PR1 and PR2 in favor of a diagnostic indication of a class 1 (C1) and class 2 (C2), respectively, for each IM associated with any combination of all OC pairs; (7) Calculate the modified diagnostic index (MDI) given by Eq. (10); (8) If $MDI \geq 0$, it is a lesion of class 1 (C1).
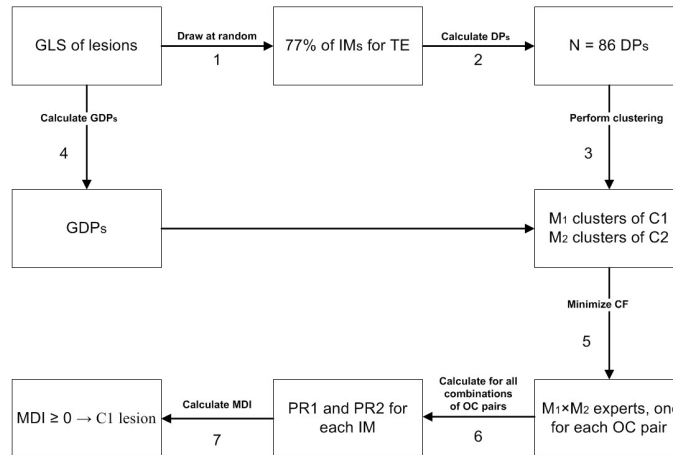

Fig. 11. Flow chart of the steps followed to obtain a diagnostic indication by a team of experts.

The decision based on the modified diagnostic index in Eq. (10) constitutes a *nonlinear* binary classifier. Its use is illustrated in Figs. 13 and 14. However, it is *not* used in the construction of the final diagnostic indication tool discussed in Section 9.

9. Final diagnostic indication tool

Consider a large set of independent lesion measurements for which the diagnosis of each lesion is known. We construct a final diagnostic indication tool by (i) drawing *at random* a training ensemble consisting of a major part (e.g. 77%) of the independent measurements on lesions belonging to each of class 1 and class 2, letting the remaining independent lesion measurements constitute a validation ensemble, and performing clustering, as discussed in Section 5.2. (*All* multiple measurements on any lesion are included into either the training or the validation ensemble.); (ii) repeating this procedure to obtain K different randomly drawn training and validation ensembles. (For randomly drawn training ensemble #k ($k = 1, 2, \ldots, K$) there will be $M_1(k) \times M_2(k) = M(k)$ different experts, each between a pair of opposite clusters.); and (iii)

constructing statistically independent experts by (a) using the set of all different experts $\mathbf{e}_{m(k)}$ $(m(k) = 1, 2, \ldots, M(k), k = 1, 2, \ldots K)$ resulting from K randomly drawn training ensembles, and the corresponding derivatives $dr^{(1)}_{m(k)}/dD$ of the partial reliabilities to compute a matrix $\mathbf{S}$ representing the covariance matrix of experts weighted with the corresponding subsets of lesion measurements of class 1, given by

$$\mathbf{S} = \sum_{k=1}^{K} \sum_{m=1}^{M(k)} \left[\int_0^{D_{max}} (dr^{(1)}_{m(k)}/dD) dD \right]^2 \mathbf{e}_{m(k)} \mathbf{e}^T_{m(k)}; \tag{11}$$

(b) finding the eigenvectors $\mathbf{s}_\lambda$ $(\lambda = 1, 2, \ldots, M^*)$ of $\mathbf{S}$; and (c) selecting, for each value of λ, those three experts that have the largest values of the scalar product $\mathbf{e}_{m(k)} \cdot \mathbf{s}_\lambda$ to obtain $3 \times M^*$ statistically independent candidate experts, and hence 3^{M^*} possible combinations of experts for the final diagnostic indication tool; and (iv) finding the best combination of experts for the final diagnostic indication tool is obtained by considering the matrices $\sum_{\lambda=1}^{M^*} \mathbf{e}_{k',\lambda} \mathbf{e}^T_{k',\lambda}$, where $\mathbf{e}_{k',\lambda}$ represents one of the 3^{M^*} possible combinations of experts, and choosing that particular combination $\mathbf{e}_{k',\lambda}$ of M^* statistically independent experts among the 3^{M^*} possible combinations, which gives the M^* largest values for the determinant of these matrices. A typical number of "best" experts is $M^* = 12$.

9.1. Application of the final diagnostic indication tool

Application of the final diagnostic indication tool described above to an unknown lesion involves: (i) calculating for each of multiple measurements (typically three) on an unknown lesion and for each of its M^* "best" experts the optimal generalized diagnostic index given by Eq. (9), finding the corresponding reliability values for class 1 and class 2, as discussed in section 7, and determining the mean value of the sum of the four largest values of the difference between the reliability for class 1 and that for class 2; and (ii) regarding the indication to be that of a class 1 lesion if the mean value determined above is greater than zero.

Experience has shown that use of this mean value provides a good combination of sensitivity and specificity.

10. Discussion

Our classification scheme was developed and optimized on a clinical data set consisting of 1,495 lesion images collected in several clinical studies using different OTD prototypes from 2005 to date. Measurements on a total of 712 lesions (including 80 malignant lesions) were used in $K = 144$ different training and validation exercises, in each of which 77% of all available measurements were randomly drawn for training, while the remaining 23% were used for validation. For these 712 lesions, the histopathological diagnoses for both the dermoscopically suspicious lesions and the dermoscopically benign lesions are given in Fig. 12. Typically three measurements were performed on each lesion [19].

An essential ingredient of our classification scheme is the construction of clusters, M_1 and M_2 for class 1 (malignant) and class 2 (benign) lesions, respectively). Each of the randomly drawn $K = 144$ training and validation ensembles provides $M_1 \times M_2$ experts (linear binary classifiers, each between a pair of opposite clusters), from which a nonlinear binary classifier or modified diagnostic index was constructed that can be used to discriminate between malignant and benign lesions. In total $M_1 \times M_2 \times 144$ different experts are provided, from which a final diagnostic indication was constructed based on a small number (typically 12) of statistically independent "best" experts. As an example, if there were $M_1 = 6$ clusters of class 1 and $M_2 = 5$ of class 2, the total number of experts would be 4,320, among which, only about 12 "best" experts would be used for construction of the final diagnostic indication tool.

Type of lesion	No. of lesions
Melanoma (n = 64)	
in situ	21
invasive	43
Basal cell carcinoma (n = 13)	
pigmented basal cell carcinoma	5
reticulated basal cell carcinoma	1
superficial basal cell carcinoma	3
nodular basal cell carcinoma	4
Squamous cell carcinoma (n = 3)	
in situ	2
invasive	1
Nevus (n = 604)	
ordinary compound	551
inflamed compound	3
irritated compound	3
congenital compound	1
lentiginous compound	1
traumatized compound	1
mildly dysplastic melanocytic	1
spitz	1
ordinary junctional	28
lentiginous junctional	2
dermal	9
intradermal	3
Keratosis (n = 28)	
porokeratosis	2
pigmented actinic keratosis	9
pigmented seborrheic keratosis	15
lichenoid actinic keratosis	2

Fig. 12. Diagnoses of lesions used for training and validation.

The accuracy A of a binary classifier is a measure of how correctly it can identify elements belonging to each of two classes, i.e.

$$A = \frac{\text{number of correct assessments}}{\text{number of all assessments}}. \tag{12}$$

Alternatively, the accuracy can be expressed in terms of the sensitivity and specifcity and the occurrence rates of the elements belonging to the two classes. If the occurrence rate is N_1 for class 1 and N_2 for class 2, and a binary classifier has sensitivity Se and specificity Sp, a measure of the accuracy is given by

$$A = \frac{N_1 \text{Se} + N_2 \text{Sp}}{N_1 + N_2}. \tag{13}$$

Figure 13 shows the performance in terms of sensitivity and specificity of our nonlinear binary classifier based on the modified diagnostic index in Eq. (10) for discriminating between malignant and benign lesions for 144 different, randomly drawn training and validation ensembles. In each of the 144 cases included in Fig. 13, the data set consisted of 34 malignant lesions and 103 benign lesions, all taken from a set of lesions considered by experienced dermatologists to be suspicious and therefore biopsied.

From Fig. 13, the sensitivity and specificity are estimated to be 0.95 and 0.20, respectively, so that Eq. (13) gives

$$A = \frac{34 \times 0.20 + 103 \times 0.95}{34 + 103} = 0.39. \tag{14}$$

In the US, around 2.5-3 million skin lesions are biopsied annually and a fraction of these – between 50,000 and 100,000 – are diagnosed as melanoma [4], implying that according to
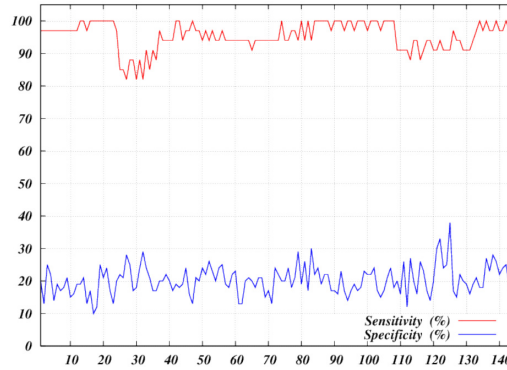
Fig. 13. Sensitivity (red) and specificity (blue) for 144 different randomly drawn training and validation sets, each consisting of 137 lesions considered to be suspicious. According to biopsy results, 34 lesions were malignant and 103 were benign.

Eq. (12), the accuracy is less than 0.04:

$$A < \frac{100,000}{2,500,000} = 0.04. \tag{15}$$

In comparison, our binary classifiers for a similar sampling of lesions would give an accuracy, according to Eq. (13), of

$$A = \frac{100,000 \times 0.95 + (2,500,000 - 100,000) \times 0.20}{2,500,000} = 0.23 \tag{16}$$

in spite of no access to medical case histories, which are available to dermatologists. Note also that our final diagnostic indication tool (see Section 9), which is based on a small number (typically 12) of "best" experts among the $M_1 \times M_2 \times 144$ binary classifiers for randomly chosen training and validation sets, gave a sensitivity higher than 0.98 for a specificity of 0.36 when applied to a set of clinically suspicious lesions [19]. This result implies that our final diagnostic indication tool can serve as a well-qualified expert, acting in a fast automatic mode to help dermatologists to arrive at the correct decision for complicated cases, and thus help eliminate unnecessary biopsies.

Figure 14 shows the performance of our binary classifiers for discriminating between malignant and benign lesions in a set of unselected ("all-comer") lesions, similar to that a Primary Care Provider (PCP) is faced with. In this case, each training and validation set includes 34 malignant lesions (as confirmed by biopsy) and 233 beign lesions (as confirmed by dermoscopy). Again the nonlinear binary classifier based on the modified diagnostic index in Eq. (10) was employed. From Fig. 14, the sensitivity and specificity are estimated to be 0.95 and 0.85, respectively, so that Eq. (13) gives

$$A = \frac{34 \times 0.95 + 233 \times 0.85}{34 + 233} = 0.86. \tag{17}$$

Since the occurrence rate of malignant lesions in an all-comer study is very low, the accuracy of our classifier is expected to be close to the value above of 0.86, which is much higher than the accuracy of a PCP. Thus, our final diagnostic indication tool can be considered as capable of providing a PCP with reliable, real-time decisions regarding melanoma referrals.
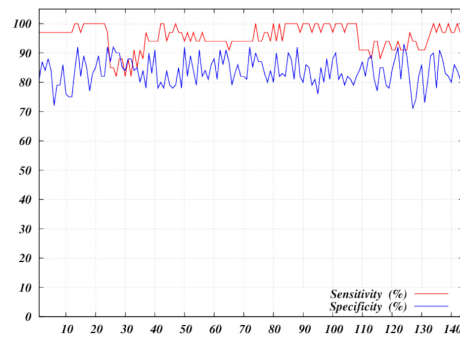
Fig. 14. Sensitivity (red) and specificity (blue) for 144 different randomly drawn training and validation sets, each consisting of 267 lesions of which 34 were malignant (according to biopsy), and 233 were benign (according to dermoscopy).

11.   Assessment of performance

The final diagnostic indication tool constructed as described in Section 9 was applied to two different sets of data including only measurements on lesions confirmed by experienced dermatologists to be benign (and therefore not biopsied) and measurements on suspicious lesions that were biopsied and diagnosed by pathologists to be either malignant or benign [19]:

- A small data set (157 lesions; 35 malignant and 39 benign, confirmed by biopsy; 81 benign, confirmed by dermoscopy) included all measurements performed on lesions that were not included in the training of the small number (typically 12) of "best" experts used for construction of the final diagnostic indication tool. The advantage of this data set is that it only contains lesions that were not used in the training of the small number of "best" experts used for construction of the final diagnostic indication tool. The disadvantage is that it is small from a statistical point of view.

- A large data set (712 lesions; 80 malignant and 217 benign, confirmed by biopsy; 415 benign, confirmed by dermoscopy) included all available measurements used for construction of the final diagnostic indication tool. The advantage of this data set is its size from a statistical point of view. A disadvantage is that it contains data that were used in the training, but the involvement in the training is not expected to have a significant impact because experience has shown that the lesion being evaluated is usually not involved in the training of the four experts taking part in the final diagnostic indication decision (see Section 9).

For discrimination of malignant lesions (as confirmed by pathology) from benign lesions (as confirmed by dermoscopy), the sensitivity/specificity was 97%/97% for application of the final diagnostic indication tool to the small data set, and 99%/93% for application of the final diagnostic indication tool to the large data set. We see that these results, obtained using the final diagnostic indication tool based on a small number of statistically independent "best" experts, are greatly improved compared to those presented in Fig. 13, obtained using the modified diagnostic index in Eq. (10).

12.   Conclusions

We have discussed a novel spectral radiometer designed to record multispectral images of skin tissue in such a way as to allow a one-dimensional configuration for a layered tissue and provide

maps of physiology properties and morphology parameters that can be used for tissue classification. Although these maps have not been verified by separate methods for measuring them, they have proved useful in an intermediate step towards a final diagnostic indication. Further, we have discussed a classification scheme based on clustering, which has been trained and validated on 712 images of pigmented skin lesions of which 415 were of benign lesions and 297 of lesions suspicious of melanoma according to clinical assessment.

The discussion in Section 10 indicates that our final dignostic indication tool can serve in a fast, automatic mode to help dermatologists discriminate between benign pigmented lesions and melanoma in a population of suspicious lesions, and provide Primary Care Providers with reliable, real-time decisions regarding melanoma referrals, thus reducing the number of unnecessary biopsies significantly. The performance assessment in Section 11 shows that application of our final diagnostic indication tool to discriminate malignant lesions (as confirmed by pathology) from benign lesions (as confirmed by dermoscopy), gives sensitivity/specificity in the range from 97%/97% (for a small data set) to 99%/93% (for a large data set).

In comparison, for discrimination between melanoma and benign pigmented lesions in a population of suspicious lesions, MelaFind conducted a clinical study involving 1,831 pigmented skin lesions, obtaining 98% sensitivity and 9% specificity [8], and Nevisense (SciBase AB, Stockholm, Sweden) employed a non-optical technique called electrical impedance spectroscopy in a clinical study involving 1,943 suspicious lesions, obtaining 96.6% sensitivity and 34.4% specificity for melanoma detection [20].

In conclusion, the optical device and associated automatic diagnostic indication tool discussed in this paper may be used to obtain a significant reduction of unnecessary, costly biopsies in melanoma detection, both by dermatologists and Primary Care Providers, without compromising patient safety.

Appendix A

In order to increase the robustness of the maps of physiology properties and morphology parameters obtained by the OTD inversion procedure, we employ statistical information extracted from multiple measurements (typically three) of each lesion. For each of the 10 different wavelengths λ_i ($i = 1, 2, \ldots, 10$), we compute the average value $I_{\lambda_i,m}$ for measurement #m of the reflected light for each pixel (after compression) inside the area surrounding the lesion. Next, we average over several measurements of the same lesion to obtain I_{λ_i}, and we define the column vectors

$$\mathbf{I}_m = [I_{\lambda_1,m}, I_{\lambda_2,m}, \ldots, \ldots I_{\lambda_{10},m}]^T \quad ; \quad \mathbf{I} = [I_{\lambda_1}, I_{\lambda_2}, \ldots, \ldots I_{\lambda_{10}}]^T \tag{18}$$

where T denotes the transpose. Further, we define the vector $\Delta\mathbf{I}_m = \mathbf{I}_m - \mathbf{I}$ and estimate the covariance matrix $\hat{\mathbf{C}}_\ell$ for lesion #ℓ:

$$\hat{\mathbf{C}}_\ell = \frac{1}{M-1} \sum_{m=1}^{M} \Delta\mathbf{I}_m [\Delta\mathbf{I}_m]^T \tag{19}$$

which is a 30×30 matrix. Finally, we average over all available lesions ($\ell = 1, 2, \ldots, L$) to obtain

$$\hat{\mathbf{C}} = \frac{1}{L} \sum_{\ell=1}^{L} \hat{\mathbf{C}}_\ell \tag{20}$$

which represents the uncertainties in our measurements.

An estimate of the misfit between the measured and simulated reflected light for a given pixel (after compression) is given by (see Eq. 9(A.7) in [17]):

$$\mathcal{J}_0 = [\mathbf{i}_{\lambda_i,n} - \tilde{\mathbf{i}}_{\lambda_i,n}]^T [\hat{\mathbf{C}}]^{-1} [\mathbf{i}_{\lambda_i,n} - \tilde{\mathbf{i}}_{\lambda_i,n}] \tag{21}$$

where $\mathbf{i}_{\lambda_i,n}$ is the measured reflected light for pixel #n and wavelength λ_i, and $\tilde{\mathbf{i}}_{\lambda_i,n}$ is the simulated reflected light for pixel #n and wavelength λ_i. Use of the result in Eq. (21) in the inversion procedure reduces the difference between the resulting maps of physiology properties and morphology parameters significantly, and results in reduced variance of the diagnostic parameters among multiple measurements on the same lesion.

Appendix B

To construct clusters of independent measurements on class 1 lesions relative to the entire set of independent measurements on class 2 lesions, we start with the class 1 measurement having the highest clustering index, as given by Eq. (6). Then we add c_m independent measurements to this one to obtain a total of $c_m + 1$ independent measurements in this first cluster, where c_m is obtained from the requirement that the function $F(c)$, given by

$$F(c) = S(c) \times \tanh(10\frac{c+1}{C}) \tag{22}$$

shall have its maximum value when $c = c_m$. Here C is the total number of independent measurements belonging to the available set of independent measurements on class 1 lesions, and $S(c)$ is the specificity, i.e. the ratio between the number of correctly classified independent measurements on class 2 lesions and the total number of independent measurements on class 2 lesions.

Similarly to the procedure for construction of class 1 clusters, we start the construction of class 2 clusters by considering the independent measurement on class 2 lesions having the highest clustering index, as defined in Eq. (6), and add c_m independent measurements to this independent measurement to obtain a total of $c_m + 1$ measurements in this *first* class 2 cluster, where c_m is obtained from the requirement that the function $F(c)$, given by Eq. (22), shall have its maximum value when $c = c_m$. Here C is the total number of measurements belonging to the available set of independent measurements on class 2 lesions, and $S(c)$ is the sum of specificities for the cluster under construction vs. each of the class 1 clusters. The specificity $S_i(c)$ for the cluster under construction vs. cluster #i of class 1 lesions is the ratio between the number of correctly classified measurements of class 2 lesions and the total number of measurements on class 2 lesions contained in the cluster under construction. Thus, $S(c)$ is given by

$$S(c) = \sum_{i=1}^{L_1} S_i(c). \tag{23}$$

Appendix C

In order to reduce the dimension of the optimization problem discussed in Section 5, we define a set of *generalized* diagnostic parameters by introducing a covariance matrix for each of the two classes ($q = 1$ and $q = 2$) of lesions, given by

$$\hat{\mathbf{V}}^{(q)} = \frac{1}{M_q - 1} \sum_{i=1}^{M_q} (\mathbf{p}_i^{(q)} - \langle \mathbf{p}^{(q)} \rangle)(\mathbf{p}_i^{(q)} - \langle \mathbf{p}^{(q)} \rangle)^T \tag{24}$$

where the superscript T denotes the transpose, $\mathbf{p}_i^{(q)} = \{p_{i,1}^{(q)}, p_{i,2}^{(q)}, \ldots, p_{i,N}^{(q)}\}$ is the vector of diagnostic parameters comprised of N components for independent measurement #i on lesions belonging to class q, $\langle \mathbf{p}^{(q)} \rangle$ is the average value of the diagnostic vectors for all independent measurements on lesions belonging to class q, and M_q is the number of independent measurements on lesions belonging to class q. Note that by definition, $\mathrm{Tr}\{\hat{\mathbf{V}}^{(1)}\} > \mathrm{Tr}\{\hat{\mathbf{V}}^{(2)}\}$. Next, we introduce a discriminating operator

$$\hat{\mathbf{D}} = [\hat{\mathbf{V}}^{(2)}]^{-1/2}[\hat{\mathbf{V}}^{(1)}][\hat{\mathbf{V}}^{(2)}]^{-1/2} \tag{25}$$

which is a generalization of the signal-to-noise ratio for multivariate random signals. To discriminate between independent measurements on lesions belonging to class 1 and class 2 and reduce the dimension of the optimization problem we extract eigenvectors $\mathbf{d}_\alpha$ according to

$$\hat{\mathbf{D}}\mathbf{d}_\alpha = \alpha\mathbf{d}_\alpha \qquad (26)$$

and introduce a subset of eigenvectors $\mathbf{d}_{\alpha_k}$ for each of which the eigenvalue $\alpha_k > \alpha_{\min}$, where $\alpha_{\min}$ is chosen to be equal to or larger than 0.7 in order to ensure that one accounts for sufficiently large variations of the diagnostic parameters associated with independent measurements on class 1 lesions. For an independent measurement on a lesion of class 1 or class 2, we define a vector $\tilde{\mathbf{p}}$ of *generalized* diagnostic parameters:

$$\tilde{\mathbf{p}} = \{\mathbf{d}_{\alpha_1} \cdot \Delta\mathbf{p}, \mathbf{d}_{\alpha_2} \cdot \Delta\mathbf{p}, \ldots, \mathbf{d}_{\alpha_{\min}} \cdot \Delta\mathbf{p}\}. \qquad (27)$$

Here $\Delta\mathbf{p} = \mathbf{p} - \langle\mathbf{p}^{(2)}\rangle$, where $\mathbf{p}$ is the vector of original diagnostic parameters in Eq. (3) and $\langle\mathbf{p}^{(2)}\rangle$ is the average of the vectors of diagnostic parameters for *all* independent measurements on class 2 lesions.

The condition $\alpha_k > \alpha_{\min}$ leads to a substantial reduction in the number of diagnostic parameters. Thus, the number $\tilde{N}$ of *generalized* diagnostic parameters is typically only one third of the number $N = 86$ of original diagnostic parameters. The *generalized* diagnostic index for an independent measurement on a lesion of class 1 or class 2 is given by:

$$D = \tilde{\mathbf{w}} \cdot \tilde{\mathbf{p}} \qquad (28)$$

where $\tilde{\mathbf{w}}$ has the same dimension as $\tilde{\mathbf{p}}$.

Conflict of interest

J.J. Stamnes is involved in Balter Medical AS as CEO (part time), board member, co-inventor of patents, and stock owner. G. Ryzhikov, M. Biryulina, B. Hamre, and L. Zhao are involved in Balter Medical AS as employees (part time), co-inventors of patents, and minor stock owners. K. Stamnes is involved in Balter Medical AS as consultant, board member, co-inventor of patents, and stock owner.